



OIFE

RAW SEQUENCE LISTING

DATE: 04/01/2002

PATENT APPLICATION: US/10/028,780

TIME: 15:16:49

Input Set : N:\Crf3\RULE60\10028780.raw

Output Set: N:\CRF3\04012002\J028780.raw

SEQUENCE LISTING

3 (1) GENERAL INFORMATION:

5 (i) APPLICANT: DUAN, D. ROXANNE

6 SHILATIFARD, ALI

7 CONAWAY, JOAN W.

8 CONAWAY, RONALD C.

10 (ii) TITLE OF INVENTION: ELL2, A New Member of an ELL Family of

11 RNA Polymerase II Elongation Factors

13 (iii) NUMBER OF SEQUENCES: 34

15 (iv) CORRESPONDENCE ADDRESS:

16 (A) ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.

17 (B) STREET: 1100 NEW YORK AVENUE, SUITE 600

18 (C) CITY: WASHINGTON

19 (D) STATE: D.C.

20 (E) COUNTRY: USA

21 (F) ZIP: 20005-3934

23 (v) COMPUTER READABLE FORM:

24 (A) MEDIUM TYPE: Floppy disk

25 (B) COMPUTER: IBM PC compatible

26 (C) OPERATING SYSTEM: PC-DOS/MS-DOS

27 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

29 (vi) CURRENT APPLICATION DATA:

C--> 30 (A) APPLICATION NUMBER: US/10/028,780

C--> 31 (B) FILING DATE: 28-Dec-2001

32 (C) CLASSIFICATION:

34 (vii) PRIOR APPLICATION DATA:

36 (A) APPLICATION NUMBER: US/09/026,343

37 (B) FILING DATE:

39 (A) APPLICATION NUMBER: US 60/038,447

40 (B) FILING DATE: 19-FEB-1997

42 (viii) ATTORNEY/AGENT INFORMATION:

43 (A) NAME: GOLDSTEIN, JORGE A.

44 (B) REGISTRATION NUMBER: 29,021

45 (C) REFERENCE/DOCKET NUMBER: 1488.0880001

47 (ix) TELECOMMUNICATION INFORMATION:

48 (A) TELEPHONE: (202) 371-2600

49 (B) TELEFAX: (202) 372-2540

52 (2) INFORMATION FOR SEQ ID NO: 1:

54 (i) SEQUENCE CHARACTERISTICS:

55 (A) LENGTH: 2139 base pairs

56 (B) TYPE: nucleic acid

57 (C) STRANDEDNESS: double

58 (D) TOPOLOGY: both

ENTERED

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/028,780

DATE: 04/01/2002

TIME: 15:16:49

Input Set : N:\Crif3\RULE60\10028780.raw

Output Set: N:\CRF3\04012002\J028780.raw

```

60      (ii) MOLECULE TYPE: cDNA
63      (ix) FEATURE:
64          (A) NAME/KEY: CDS
65          (B) LOCATION: 94..2013
68      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
70 CAGTGGCGGC GGGTGCAGAA GCCCAAGCAG CGCGGCCGCA GTGGAGGCTA GAGCCGGAGC      60
72 GGCGGCGGCG GCGGCACCCC GGGGAGGTTT AAG ATG GCG GCG GGG GGG ACA GGG      114
73                                     Met Ala Ala Gly Gly Thr Gly
74                                     1           5
76 GGC CTG CGG GAG GAG CAG CGC TAT GGG CTG TCG TGC GGA CGG CTG GGG      162
77 Gly Leu Arg Glu Glu Gln Arg Tyr Gly Leu Ser Cys Gly Arg Leu Gly
78      10           15           20
80 CAG GAC AAC ATC ACC GTA CTG CAT GTG AAG CTC ACC GAG ACG GCG ATC      210
81 Gln Asp Asn Ile Thr Val Leu His Val Lys Leu Thr Glu Thr Ala Ile
82      25           30           35
84 CGG GCG CTC GAG ACT TAC CAG AGC CAC AAG AAT TTA ATT CCT TTT CGA      258
85 Arg Ala Leu Glu Thr Tyr Gln Ser His Lys Asn Leu Ile Pro Phe Arg
86      40           45           50           55
88 CCT TCA ATC CAG TTC CAA GGA CTC CAC GGG CTT GTC AAA ATT CCC AAA      306
89 Pro Ser Ile Gln Phe Gln Gly Leu His Gly Leu Val Lys Ile Pro Lys
90      60           65           70
92 AAT GAT CCC CTC AAT GAA GTT CAT AAC TTT AAC TTT TAT TTG TCA AAT      354
93 Asn Asp Pro Leu Asn Glu Val His Asn Phe Asn Phe Tyr Leu Ser Asn
94      75           80           85
96 GTG GGC AAA GAC AAC CCT CAG GGC AGC TTT GAC TGC ATC CAG CAA ACA      402
97 Val Gly Lys Asp Asn Pro Gln Gly Ser Phe Asp Cys Ile Gln Gln Thr
98      90           95           100
100 TTC TCC AGC TCT GGA GCC TCC CAG CTC AAT TGC CTG GGA TTT ATA CAA      450
101 Phe Ser Ser Ser Gly Ala Ser Gln Leu Asn Cys Leu Gly Phe Ile Gln
102      105           110           115
104 GAT AAA ATT ACA GTG TGT GCA ACA AAC GAC TCG TAT CAG ATG ACA CGA      498
105 Asp Lys Ile Thr Val Cys Ala Thr Asn Asp Ser Tyr Gln Met Thr Arg
106      120           125           130           135
108 GAA AGA ATG ACC CAG GCA GAG GAG GAA TCC CGC AAC CGA AGC ACA AAA      546
109 Glu Arg Met Thr Gln Ala Glu Glu Glu Ser Arg Asn Arg Ser Thr Lys
110      140           145           150
112 GTT ATC AAA CCC GGT GGA CCA TAT GTA GGG AAA AGA GTG CAA ATT CGG      594
113 Val Ile Lys Pro Gly Gly Pro Tyr Val Gly Lys Arg Val Gln Ile Arg
114      155           160           165
116 AAA GCA CCT CAA GCT GTT TCA GAT ACA GTT CCT GAG AGG AAA AGG TCA      642
117 Lys Ala Pro Gln Ala Val Ser Asp Thr Val Pro Glu Arg Lys Arg Ser
118      170           175           180
120 ACC CCC ATG AAC CCT GCA AAT ACA ATT CGA AAG ACA CAT AGC AGC AGC      690
121 Thr Pro Met Asn Pro Ala Asn Thr Ile Arg Lys Thr His Ser Ser Ser
122      185           190           195
124 ACC ATC TCT CAG AGG CCA TAC AGG GAC AGG GTG ATT CAC TTA CTG GCC      738
125 Thr Ile Ser Gln Arg Pro Tyr Arg Asp Arg Val Ile His Leu Leu Ala
126      200           205           210           215
128 CTG AAG GCC TAC AAG AAA CCG GAG CTA CTT GCT AGA CTC CAG AAA GAT      786

```

RAW SEQUENCE LISTING

DATE: 04/01/2002

PATENT APPLICATION: US/10/028,780

TIME: 15:16:49

Input Set : N:\Crif3\RULE60\10028780.raw

Output Set: N:\CRF3\04012002\J028780.raw

129	Leu	Lys	Ala	Tyr	Lys	Lys	Pro	Glu	Leu	Leu	Ala	Arg	Leu	Gln	Lys	Asp	
130					220					225					230		
132	GGT	GTC	AAT	CAA	AAA	GAC	AAG	AAC	TCC	CTG	GGA	GCA	ATT	CTG	CAA	CAG	834
133	Gly	Val	Asn	Gln	Lys	Asp	Lys	Asn	Ser	Leu	Gly	Ala	Ile	Leu	Gln	Gln	
134				235					240					245			
136	GTA	GCC	AAT	CTG	AAT	TCT	AAG	GAC	CTC	TCA	TAT	ACC	TTA	AAG	GAT	TAT	882
137	Val	Ala	Asn	Leu	Asn	Ser	Lys	Asp	Leu	Ser	Tyr	Thr	Leu	Lys	Asp	Tyr	
138				250					255					260			
140	GTT	TTT	AAA	GAG	CTT	CAA	AGA	GAC	TGG	CCT	GGA	TAC	AGT	GAA	ATA	GAC	930
141	Val	Phe	Lys	Glu	Leu	Gln	Arg	Asp	Trp	Pro	Gly	Tyr	Ser	Glu	Ile	Asp	
142				265					270					275			
144	AGA	CGG	TCA	TTG	GAG	TCA	GTG	CTC	TCT	AGA	AAA	CTA	AAT	CCG	TCT	CAG	978
145	Arg	Arg	Ser	Leu	Glu	Ser	Val	Leu	Ser	Arg	Lys	Leu	Asn	Pro	Ser	Gln	
146	280						285					290				295	
148	AAT	GCT	ACA	GGC	ACC	AGC	CGT	TCA	GAA	TCT	CCT	GTA	TGT	TCT	AGT	AGA	1026
149	Asn	Ala	Thr	Gly	Thr	Ser	Arg	Ser	Glu	Ser	Pro	Val	Cys	Ser	Ser	Arg	
150					300						305				310		
152	GAT	GCT	GTA	TCT	TCT	CCT	CAG	AAA	CGG	CTT	TTG	GAT	TCA	GAG	TTT	ATT	1074
153	Asp	Ala	Val	Ser	Ser	Pro	Gln	Lys	Arg	Leu	Leu	Asp	Ser	Glu	Phe	Ile	
154				315					320					325			
156	GAT	CCT	TTA	ATG	AAT	AAA	AAA	GCC	CGA	ATA	TCT	CAC	CTG	ACG	AAC	AGA	1122
157	Asp	Pro	Leu	Met	Asn	Lys	Lys	Ala	Arg	Ile	Ser	His	Leu	Thr	Asn	Arg	
158			330						335					340			
160	GTA	CCA	CCA	ACA	CTA	AAT	GGT	CAT	TTG	AAT	CCC	ACC	AGT	GAA	AAA	TCG	1170
161	Val	Pro	Pro	Thr	Leu	Asn	Gly	His	Leu	Asn	Pro	Thr	Ser	Glu	Lys	Ser	
162			345					350					355				
164	GCT	GCA	GGC	CTC	CCA	CTG	CCC	CCT	GCG	GCT	GCT	GCC	ATC	CCC	ACC	CCT	1218
165	Ala	Ala	Gly	Leu	Pro	Leu	Pro	Pro	Ala	Ala	Ala	Ala	Ile	Pro	Thr	Pro	
166	360					365						370				375	
168	CCA	CCG	CTG	CCT	TCA	ACC	TAT	CTG	CCC	ATC	TCA	CAT	CCT	CCT	CAG	ATT	1266
169	Pro	Pro	Leu	Pro	Ser	Thr	Tyr	Leu	Pro	Ile	Ser	His	Pro	Pro	Gln	Ile	
170					380						385				390		
172	GTA	AAT	TCT	AAC	TCC	AAC	TCC	CCT	AGC	ACT	CCA	GAA	GGC	CGG	GGG	ACT	1314
173	Val	Asn	Ser	Asn	Ser	Asn	Ser	Pro	Ser	Thr	Pro	Glu	Gly	Arg	Gly	Thr	
174				395					400					405			
176	CAA	GAC	CTA	CCT	GTT	GAC	AGT	TTT	AGT	CAA	AAC	GAT	AGT	ATC	TAT	GAG	1362
177	Gln	Asp	Leu	Pro	Val	Asp	Ser	Phe	Ser	Gln	Asn	Asp	Ser	Ile	Tyr	Glu	
178			410						415					420			
180	GAC	CAG	CAA	GAC	AAA	TAT	ACC	TCT	AGG	ACT	TCT	CTG	GAA	ACC	TTA	CCC	1410
181	Asp	Gln	Gln	Asp	Lys	Tyr	Thr	Ser	Arg	Thr	Ser	Leu	Glu	Thr	Leu	Pro	
182			425					430					435				
184	CCT	GGT	TCC	GTT	CTA	CTA	AAG	TGT	CCA	AAG	CCT	ATG	GAA	GAA	AAC	CAT	1458
185	Pro	Gly	Ser	Val	Leu	Leu	Lys	Cys	Pro	Lys	Pro	Met	Glu	Glu	Asn	His	
186	440					445						450				455	
188	TCA	ATG	TCT	CAC	AAA	AAG	TCC	AAA	AAG	AAG	TCT	AAA	AAA	CAT	AAG	GAA	1506
189	Ser	Met	Ser	His	Lys	Lys	Ser	Lys	Lys	Lys	Ser	Lys	Lys	His	Lys	Glu	
190					460						465				470		
192	AAG	GAC	CAA	ATA	AAA	AAG	CAC	GAC	ATT	GAG	ACT	ATT	GAG	GAA	AAG	GAG	1554
193	Lys	Asp	Gln	Ile	Lys	Lys	His	Asp	Ile	Glu	Thr	Ile	Glu	Glu	Lys	Glu	

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/028,780

DATE: 04/01/2002

TIME: 15:16:50

Input Set : N:\Crif3\RULE60\10028780.raw

Output Set: N:\CRF3\04012002\J028780.raw

```

194          475          480          485
196 GAA GAT CTT AAG AGA GAA GAG GAA ATT GCC AAG CTA AAT AAC TCC AGT      1602
197 Glu Asp Leu Lys Arg Glu Glu Glu Ile Ala Lys Leu Asn Asn Ser Ser
198          490          495          500
200 CCA AAT TCC AGT GGA GGA GTT AAA GAG GAT TGC ACT GCC TCC ATG GAA      1650
201 Pro Asn Ser Ser Gly Gly Val Lys Glu Asp Cys Thr Ala Ser Met Glu
202          505          510          515
204 CCT TCA GCA ATT GAA CTC CCA GAT TAT TTG ATA AAA TAT ATC GCT ATC      1698
205 Pro Ser Ala Ile Glu Leu Pro Asp Tyr Leu Ile Lys Tyr Ile Ala Ile
206 520          525          530          535
208 GTC TCC TAT GAG CAA CGC CAG AAT TAT AAG GAT GAC TTC AAT GCA GAG      1746
209 Val Ser Tyr Glu Gln Arg Gln Asn Tyr Lys Asp Asp Phe Asn Ala Glu
210          540          545          550
212 TAT GAT GAG TAC AGA GCT TTG CAT GCC AGG ATG GAG ACT GTA GCT AGA      1794
213 Tyr Asp Glu Tyr Arg Ala Leu His Ala Arg Met Glu Thr Val Ala Arg
214          555          560          565
216 AGA TTT ATC AAA CTA GAT GCA CAA AGA AAG CGC CTT TCT CCA GGC TCA      1842
217 Arg Phe Ile Lys Leu Asp Ala Gln Arg Lys Arg Leu Ser Pro Gly Ser
218          570          575          580
220 AAA GAG TAT CAG AAT GTT CAT GAA GAA GTC TTA CAA GAA TAT CAG AAG      1890
221 Lys Glu Tyr Gln Asn Val His Glu Glu Val Leu Gln Glu Tyr Gln Lys
222          585          590          595
224 ATA AAG CAG TCT AGT CCC AAT TAC CAT GAA GAA AAA TAC AGA TGT GAA      1938
225 Ile Lys Gln Ser Ser Pro Asn Tyr His Glu Glu Lys Tyr Arg Cys Glu
226 600          605          610          615
228 TAT CTT CAT AAC AAG CTG GCT CAC ATC AAA AGG CTA ATA GGT GAA TTT      1986
229 Tyr Leu His Asn Lys Leu Ala His Ile Lys Arg Leu Ile Gly Glu Phe
230          620          625          630
232 GAC CAA CAG CAA GCA GAG TCA TGG TCC TAGAACTCTG CTTGGACCAG      2033
233 Asp Gln Gln Gln Ala Glu Ser Trp Ser
234          635          640
236 AAGATGTGAA TAAACTTAAG CTTATTTATT TAAAATTCCA AATGAGTTGC TCTAGATTCT      2093
238 AAAAAGGTGA AACTTTGGCT GTTGAAAGTT TCAGTATTAG TAAACT      2139
241 (2) INFORMATION FOR SEQ ID NO: 2:
243 (i) SEQUENCE CHARACTERISTICS:
244 (A) LENGTH: 640 amino acids
245 (B) TYPE: amino acid
246 (D) TOPOLOGY: linear
248 (ii) MOLECULE TYPE: protein
250 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
252 Met Ala Ala Gly Gly Thr Gly Gly Leu Arg Glu Glu Gln Arg Tyr Gly
253 1 5 10 15
255 Leu Ser Cys Gly Arg Leu Gly Gln Asp Asn Ile Thr Val Leu His Val
256 20 25 30
258 Lys Leu Thr Glu Thr Ala Ile Arg Ala Leu Glu Thr Tyr Gln Ser His
259 35 40 45
261 Lys Asn Leu Ile Pro Phe Arg Pro Ser Ile Gln Phe Gln Gly Leu His
262 50 55 60
264 Gly Leu Val Lys Ile Pro Lys Asn Asp Pro Leu Asn Glu Val His Asn

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/028,780

DATE: 04/01/2002

TIME: 15:16:50

Input Set : N:\Crif3\RULE60\10028780.raw

Output Set: N:\CRF3\04012002\J028780.raw

265	65					70						75					80
267	Phe	Asn	Phe	Tyr	Leu	Ser	Asn	Val	Gly	Lys	Asp	Asn	Pro	Gln	Gly	Ser	
268					85					90						95	
270	Phe	Asp	Cys	Ile	Gln	Gln	Thr	Phe	Ser	Ser	Ser	Gly	Ala	Ser	Gln	Leu	
271				100					105					110			
273	Asn	Cys	Leu	Gly	Phe	Ile	Gln	Asp	Lys	Ile	Thr	Val	Cys	Ala	Thr	Asn	
274			115					120					125				
276	Asp	Ser	Tyr	Gln	Met	Thr	Arg	Glu	Arg	Met	Thr	Gln	Ala	Glu	Glu	Glu	
277		130					135					140					
279	Ser	Arg	Asn	Arg	Ser	Thr	Lys	Val	Ile	Lys	Pro	Gly	Gly	Pro	Tyr	Val	
280	145					150					155					160	
282	Gly	Lys	Arg	Val	Gln	Ile	Arg	Lys	Ala	Pro	Gln	Ala	Val	Ser	Asp	Thr	
283				165						170					175		
285	Val	Pro	Glu	Arg	Lys	Arg	Ser	Thr	Pro	Met	Asn	Pro	Ala	Asn	Thr	Ile	
286			180						185					190			
288	Arg	Lys	Thr	His	Ser	Ser	Ser	Thr	Ile	Ser	Gln	Arg	Pro	Tyr	Arg	Asp	
289		195						200					205				
291	Arg	Val	Ile	His	Leu	Leu	Ala	Leu	Lys	Ala	Tyr	Lys	Lys	Pro	Glu	Leu	
292		210					215					220					
294	Leu	Ala	Arg	Leu	Gln	Lys	Asp	Gly	Val	Asn	Gln	Lys	Asp	Lys	Asn	Ser	
295	225					230					235					240	
297	Leu	Gly	Ala	Ile	Leu	Gln	Gln	Val	Ala	Asn	Leu	Asn	Ser	Lys	Asp	Leu	
298				245						250					255		
300	Ser	Tyr	Thr	Leu	Lys	Asp	Tyr	Val	Phe	Lys	Glu	Leu	Gln	Arg	Asp	Trp	
301			260						265					270			
303	Pro	Gly	Tyr	Ser	Glu	Ile	Asp	Arg	Arg	Ser	Leu	Glu	Ser	Val	Leu	Ser	
304		275						280						285			
306	Arg	Lys	Leu	Asn	Pro	Ser	Gln	Asn	Ala	Thr	Gly	Thr	Ser	Arg	Ser	Glu	
307		290					295					300					
309	Ser	Pro	Val	Cys	Ser	Ser	Arg	Asp	Ala	Val	Ser	Ser	Pro	Gln	Lys	Arg	
310	305					310					315					320	
312	Leu	Leu	Asp	Ser	Glu	Phe	Ile	Asp	Pro	Leu	Met	Asn	Lys	Lys	Ala	Arg	
313				325						330					335		
315	Ile	Ser	His	Leu	Thr	Asn	Arg	Val	Pro	Pro	Thr	Leu	Asn	Gly	His	Leu	
316			340						345					350			
318	Asn	Pro	Thr	Ser	Glu	Lys	Ser	Ala	Ala	Gly	Leu	Pro	Leu	Pro	Pro	Ala	
319		355						360						365			
321	Ala	Ala	Ala	Ile	Pro	Thr	Pro	Pro	Pro	Leu	Pro	Ser	Thr	Tyr	Leu	Pro	
322		370					375						380				
324	Ile	Ser	His	Pro	Pro	Gln	Ile	Val	Asn	Ser	Asn	Ser	Asn	Ser	Pro	Ser	
325	385					390					395					400	
327	Thr	Pro	Glu	Gly	Arg	Gly	Thr	Gln	Asp	Leu	Pro	Val	Asp	Ser	Phe	Ser	
328				405						410					415		
330	Gln	Asn	Asp	Ser	Ile	Tyr	Glu	Asp	Gln	Gln	Asp	Lys	Tyr	Thr	Ser	Arg	
331			420						425					430			
333	Thr	Ser	Leu	Glu	Thr	Leu	Pro	Pro	Gly	Ser	Val	Leu	Leu	Lys	Cys	Pro	
334		435						440						445			
336	Lys	Pro	Met	Glu	Glu	Asn	His	Ser	Met	Ser	His	Lys	Lys	Ser	Lys	Lys	
337		450					455					460					

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/028,780

DATE: 04/01/2002

TIME: 15:16:51

Input Set : N:\Crf3\RULE60\10028780.raw

Output Set: N:\CRF3\04012002\J028780.raw

L:30 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:31 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:447 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=7
L:580 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=8